UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
ENVIRONMENTAL RESPONSE TEAM CENTER
2890 WOODBRIDGE AVENUE, BUILDING 18
EDISON, NEW JERSEY 08837

DATE: April 21, 2004

TO: Jon D. Rauscher, Ph.D., U.S. EPA Region 6

FROM: Marc S. Greenberg, Ph.D., Environmental Toxicologist, U.S. EPA/ERT/WSU-IPA

SUBJECT: TECHNICAL MEMORANDUM - DERIVATION OF A COMMUNITY SPECIFIC M_{SD} FOR EL PASO, TEXAS

SUMMARY

United States Environmental Protection Agency/Emergency Response Team (U.S. EPA/ERT) evaluated the relationship of lead (Pb) concentrations in indoor dust to Pb concentrations in soil and indoor paint at residences near a lead smelter in El Paso, Texas (TX). U.S. EPA/ERT was provided with the U.S. EPA Region 6 *Standard Operating Procedure for Estimating the Soil to Dust Variable* (SOP) (U.S. EPA 2004) as a reference for this task. The variable discussed in this SOP, the mass fraction of indoor dust derived from outdoor soil (M_{SD}) is one of many inputs to the U.S. EPA Integrated Exposure Uptake Biokinetic (IEUBK) Model for predicting blood lead levels in children (U.S. EPA 1994). The objective was to investigate the derivation of a community-specific M_{SD} for use as an alternative to the default value of 0.70 in the IEUBK model for the prediction of blood lead concentrations in children in this Texas community.

This Technical Memorandum (TM) explores three methods for deriving a community-specific M_{SD} based on data acquired from 30 residential properties in the vicinity of the lead smelter plant. The 0.70 M_{SD} default value was based on empirical relationships between soil and dust Pb concentrations measured in various residential communities. The use and derivation of the M_{SD} variable is described in detail in the U.S. EPA's *Short Sheet: IEUBK Model Mass Fraction of Soil in Indoor Dust (M_{SD}) Variable* (U.S. EPA 1998), which recommends the use of actual indoor dust lead concentrations (Pb_D) as IEUBK model inputs, when available. In the absence of site-specific Pb_D data, the default M_{SD} value may be used, as described below, to derive Pb_D. A community-specific M_{SD} may be used as an alternative to the default M_{SD} value when compelling evidence suggests that the mass fraction of soil in indoor dust differs from the default. In this study of homes potentially impacted by a nearby lead smelter, data from several sampled media (indoor dust, outdoor soil, and indoor and outdoor paint), and categorical variables reflecting potential influences on M_{SD} (lead paint, vegetation, flooring type, and pets) were analyzed in an effort to come up with an alternative value for M_{SD} that would be more representative of conditions in this neighborhood.

When soil is the predominant source of lead in house dust, M_{SD} can be calculated as the ratio between Pb_D and lead concentrations in soil (Pb_S); that is, $M_{SD}= Pb_D/Pb_S$ (U.S. 1998). Solving for Pb in dust, this becomes $Pb_D=M_{SD}*Pb_S$. This equation is useful only in situations where Pb_0 (the concentration of lead in dust derived from non-soil sources) is assumed to be zero (U.S. EPA 1998). When there are other significant non-soil sources of lead, Pb_D represents the sum of the contributions from all these sources (Pb_0+Pb_S), and therefore, Pb_D will be greater than that calculated by $M_{SD}*Pb_S$. In such cases, attempts to use measured Pb concentration data for soil and dust to estimate M_{SD} become problematic (U.S. EPA 1998). The El Paso data show that lead paint is present in most of the tested homes. In 17 of these homes,

the interior paint conditions are reported to be poor (based on visual assessment by a certified lead paint inspector), suggesting a potentially significant non-soil source for lead in house dust. Therefore, attempts to derive an alternative M_{SD} using the soil and dust data from all 30 properties were unsuccessful.

A regression analysis was chosen as an alternative for estimating the ratio of the mean dust concentration to the mean soil concentration to develop a central value for M_{SD} in the community. This approach is consistent with EPA guidance on the M_{SD} variable (U.S. EPA 1998). From this analysis, we determined a scientifically valid M_{SD} of 0.57 that would provide an appropriate alternative to the default value of 0.70 in the IEUBK model for the prediction of blood lead concentrations in children in this TX community.

DATA ANALYSIS

Data were provided for 30 residential properties identified by a "Blind Property ID" , which will be referred to as the property ID throughout this TM and in corresponding statistical output. Data included Pb concentrations in milligrams per kilogram (mg/kg) in indoor dust, paint and soil for each of the 30 properties, as well as categorical information regarding the presence/absence of indoor lead paint, the presence/absence of outdoor lead paint, the condition of the indoor paint (good/poor), the presence/absence of vegetative cover in the surrounding yards, the type of flooring inside the residences (soft/hard/mixed), and the presence/absence of pets with movement between the inside and the outside of the individual residences (e.g., cats, dogs). The soil samples were taken using by uniform grid design in each yard. The sieved sample results were used in the statistical analyses (non-sieved samples may contain paint chips or fibrous organic material that may significantly bias measured lead concentrations [U.S. EPA 1998]). Multiple soil, dust, and paint samples were collected at each residence; therefore, a mean Pb concentration was calculated per media per property, and these means were used in the statistical analyses. This is in accordance with U.S. EPA's *Risk Assessment Guidance for Superfund* (U.S. EPA 1989), which calls for the use of the arithmetic average as a reasonable estimate of the concentration to which a person may be exposed over time. Statistical analyses were also conducted using the maximum media Pb concentration per property, because at residences where Pb contamination was unevenly distributed across the property, the use of the property mean could effectively mask the Pb contamination. By evaluating maximum concentrations, worst case conditions could be explored. Statistical analyses were performed using SAS® version 8.00. Detailed statistical output generated using SAS® can be found in Attachment A.

Several methods were attempted to derive a community-specific M_{SD}. These methods are outlined below.

Method 1: Empirically Derived M_{SD}

An M_{SD} was calculated for each of the 30 properties using the relationship defined in the U.S. EPA Region 6 SOP, $M_{SD} = Pb_D/Pb_S$. Mean Pb concentrations in dust were divided by corresponding mean Pb concentrations in soil to obtain property-specific M_{SD}. The median value, 0.49, was then computed for the 30 empirically derived M_{SD}. The median was chosen over the mean because of the non-normal distribution of the population of M_{SD} values. EPA guidance cites that this methodology is only representative of true Pb conditions in indoor dust if the only expected contributing factor is outdoor soil Pb (U.S. EPA 1998). In this community there is evidence that this is not the case. As stated above, all but two of the properties sampled in this study had the presence of Pb paint either inside or outside of their residences; therefore, an unknown bias exists in the calculation of these empirically derived M_{SD} values. Therefore, this process was repeated for a subset of 13 properties with indoor paint conditions categorized as good. The resulting median was 0.42.

Method 2: Linear Regression of Indoor Dust and Soil

Using the linear relationship describing the soil to dust mass fraction, $Pb_D = M_{SD} * Pb_S$, linear regression analyses were computed using soil Pb concentrations (independent variable, x), to predict indoor dust Pb concentrations (dependent variable, y)(see Equation 1, U.S. EPA 2004). Using this relationship the coefficient of the independent variable, Pb in soil, is the derived M_{SD}. If the only source of Pb in indoor dust is Pb in soil, then

M_{SD}= Pb_D/Pb_S, as noted in Equation 1, and there is no reported y-intercept as its value would be predicted to pass through the origin (0,0). However, if factors other than soil are contributing to Pb in indoor dust, the regression equation becomes Equation 2 in the SOP, $Pb_D=M_{SD}*Pb_S+Pb_0$.

Multiple iterations of linear regression analyses were implemented to examine these relationships. First, mean Pb concentrations for indoor dust and soil were used, followed by analyses of the maximum Pb concentrations of indoor dust and soil. Relationships including all 30 properties were explored, as well as subsets of properties based on the condition of the indoor paint (good/ poor). Outliers were identified through examination of residual plots, studentized residuals, and Cook's D statistic. The models were run with and without the outliers. Table 1 summarizes the results. In total, twelve linear regression models were computed. Of these twelve models only three were statistically significant (defined as a model with a probability-value [p-value] less than 0.05): 1) mean indoor dust concentrations versus mean soil concentrations for properties with paint condition categorized as good (N=13, p-value=0.0240, R^2=0.3837), 2) mean indoor dust concentrations versus mean soil concentrations for properties with paint condition categorized as good excluding property 30, which was identified as an outlier (N=12, p-value=0.0068, R^2=0.5357), and 3) maximum indoor dust concentrations versus maximum soil concentrations for properties with paint condition categorized as good, excluding property 30 which was identified as an outlier (N=12, p-value=0.0011, R^2=0.6709). Of these three cases only two models resulted in R^2>0.50. Both models were based on properties with indoor paint in good condition with property 30 removed, one utilizing mean concentrations per property and the other using maximum concentrations. The resulting regression equations were:

1) mean indoor dust Pb (mg/kg)=[0.56*mean soil Pb (mg/kg)] - 47.61

and

2) maximum indoor dust Pb (mg/kg)=[0.57*maximum soil Pb (mg/kg)] - 57.57

In both cases the derived M_{SD} was similar, with a value of 0.56 derived from the mean Pb concentrations and 0.57 derived from the maximum Pb concentrations. As seen in these equations, negative y-intercepts (-47.61 and -57.57) resulted from the regression analyses. Positive y-intercepts would support the idea that factors other than soil are contributing to Pb in dust. Based upon the significant regression analyses summarized above, this was not observed.

<u>Method 3: Stepwise Regression with Interval and Categorical Variables</u>

Because the R^2 values obtained by the linear regressions in Method 2 indicated that, for this residential community, factors other than soil were responsible for the Pb concentrations found in indoor dust, stepwise regression analysis was utilized to determine if any of the conditions represented by the categorical variables were influencing indoor dust Pb concentrations. These particular categorical variables (lead paint, vegetation, flooring type, pets) represented conditions cited in the Short Sheet (U.S. EPA 1998) and the SOP as having possible influence on indoor dust Pb concentrations.

Categorical variables that multiple response options were coded in a binary format known as "dummy variables" (Hardy 1993). Dummy variables allow categorical variables to be analyzed using stepwise regression. For example, the categorical variable "presence/absence of indoor Pb paint" was expressed as two dummy variables: "presence of indoor Pb paint" and "absence of indoor Pb paint". If a property had indoor Pb paint present it received a value of 1 for "presence of indoor Pb paint" and a 0 for "absence of indoor Pb paint". Each categorical variable was transformed into n dummy variables which represented the number of possible responses. A categorical variable such as "type of indoor floor covering" with three possible responses of soft, hard and mixed, resulted in three dummy variables (0, 1 and 2). Of those n dummy variables per categorical variable, only n–1 were included in the stepwise regression analysis. This derives from the requirement of the classical linear regression model, that no perfect collinearity may exist between independent variables, and therefore no independent variable in a regression model may be a linear combination of other independent variables (Hardy 1993, p.8). For example, in the case of the categorical variable "presence/absence of indoor

Pb paint", it is not appropriate to include both corresponding dummy variables "presence of indoor Pb paint" and "absence of indoor Pb paint" because a perfect linear relationship exists between the two dummy variables: "presence of indoor Pb paint" = 1 - "absence of indoor Pb paint" (i.e., one variable can be derived from the other). In all cases, the dummy variable excluded from the stepwise regression represented a reference condition (absence of indoor paint, absence of outdoor paint, paint in good condition, presence of vegetative cover, soft floor covering, absence of pets). For a more detailed explanation of regression analysis with dummy variables see Hardy (1993).

Stepwise regression analyses were conducted following a similar methodology to the linear regressions computed in Method 2. Stepwise regressions were computed using mean Pb concentrations per property and then again using maximum Pb concentrations per property for indoor dust, indoor paint, and outdoor soil. These regressions were also run with and without the outliers identified in Method 2. All categorical variables were represented in each of the iterations with the exception of the presence/absence of indoor Pb paint. This category was excluded because only 2 of the 30 properties did not have indoor Pb paint, making it an extraneous parameter in the model (there would be little variance in response to this variable). Additionally, when the data were broken into subsets based on indoor Pb paint condition, the corresponding dummy variables for paint condition were not included in the model, again because it became an extraneous parameter in the model (e.g., for the subset of properties with indoor paint in good condition the variable "indoor paint in good condition" would always equal 1).

Only variables that contributed to models at a significance level of 0.05 were included in the models. Of the twelve stepwise models computed, only three resulted in significant regression models. These models are: 1) mean indoor dust concentrations versus mean soil concentrations for properties with paint condition categorized as good, 2) mean indoor dust concentrations versus mean soil concentrations for properties with paint condition categorized as good excluding property 30, and 3) maximum indoor dust concentrations versus maximum soil concentrations and the presence/absence of pets, for properties with paint condition categorized as good, excluding property 30. Thus, for properties with indoor paint in good condition and with property 30 removed, the only additional influencing factor on indoor dust Pb concentration, other than soil, identified by the stepwise regressions was the presence/absence of pets.

A linear regression analysis was then computed for this specific subset of data including the variables maximum indoor dust Pb concentration, property maximum soil Pb concentrations, and presence of pets. The resulting model was significant with a p-value of 0.0008 and an R^2=0.7978. The resulting regression equation was:

maximum indoor dust Pb (mg/kg)=0.52*(maximum soil Pb [mg/kg]) + 78.04*(presence of pets) - 63.09

Based on this regression model, residences with pets have indoor dust Pb concentrations 78.04 mg/kg more than those without pets. The derived M_{SD} from this regression equation is 0.52. This subset of data is representative of the larger data set regarding presence/absence of pets (33% of the properties with lead paint in good condition own pets [n = 4], and 33% of the 30 properties, in total, own pets [n = 10]). This analysis indicates the role of pets in the resulting indoor dust Pb concentration but on its own is not useful in determining a community wide M_{SD} value.

CONCLUSIONS AND UNCERTAINTIES

As stated earlier, the extensive presence of indoor/outdoor Pb paint in the set of properties sampled would imply that factors other than Pb contamination in soil would contribute to Pb contamination in indoor dust and therefore confound the calculation of a community specific $M_{SD,}$ mass-fraction of indoor dust derived from outside soil. Both the U.S. EPA Region 6 SOP and U.S. EPA Short Sheet (U.S. EPA 1998) recognize the potential problems involved in deriving a community specific $M_{SD,}$ when there are sources other than soil contributing to the indoor dust Pb levels. Methods 2 and 3 both indicate that there are outside contributing factors. However, other than the presence/absence of pets identified in Method 3, the expected contributing factors (U.S. EPA 1998) could not be related statistically to the Pb levels found in indoor dust within the 30 houses sampled from this El Paso, TX community.

Based on recommendations from the U.S. EPA Short Sheet (U.S. EPA 1998), subsets were created from the data collected from these 30 properties based on the condition of the indoor paint (good/ poor). These subsets were evaluated using linear regression and stepwise regression analyses including categorical and measurement variables. Outliers were identified as property 30 for the subset of properties with paint in good condition and property 10 for properties with paint in poor condition. Regression analyses were run including these outliers and then again excluding these outliers. Data were evaluated using mean values of indoor dust, soil, and indoor paint Pb concentrations per property, as well as maximum values per property (to evaluate a worst case scenario). Results from evaluation of the means and the maximums under Method 2 resulted in similar M_{SD} values, 0.56 and 0.57 for properties with indoor paint categorized as being in good condition (with property 30 removed from the analyses). Table 1 summarizes these results. Regression analyses of the subset of properties with paint in good condition with the outliers removed were statistically significant (p-value<0.05), and resulted in greater R^2 values. When the presence/absence of pets was added to the regression analyses, based on results from the stepwise analysis using maximum Pb concentrations per property, the derived M_{SD} for properties with indoor paint in good condition and pets (property 30 removed) was reduced by only 0.04 or 0.05.

Although derivations of a community specific M_{SD}, based on this set of 30 properties, yielded consistent and similar results for properties with indoor paint in good condition, these results are derived from a limited data set. In addition to the limited data set, the following factors could have added unknown bias and must be taken into consideration when determining the applicability of these derived values:

- Deteriorated interior lead paint was identified as a potential source of lead in Pb_D. Limiting the calculation of M_{SD} to those homes with paint in good condition yields a small subset of homes (N=13; N=12 when outliers were removed). There is some indication that this subset of good condition paint homes does not vary from the larger group of 30 sampled homes with respect to other potential contributors to lead in dust (e.g., the percentage of homes with pets was identical for good and poor condition paint subsets. Thus, extrapolation of an M_{SD} derived from this subset of sampled homes is assumed to be representative of the larger community.

- Indoor paint and dust samples, while collected according to U.S. EPA protocol, were not necessarily collected from the same locations within each residence. This may lead to uncertainty in estimating the indoor dust concentrations of lead. However this uncertainty was reduced for by evaluating both average and maximum concentrations of lead in dust.

- There is uncertainty from lack of information on additional factors that may have contributed to lead in indoor dust. Examples of these factors include the frequency of house cleaning, method of cleaning (wet versus dry), windows kept open or closed, Pb dust carried home on work clothes, or hobbies that may involve lead exposure.

Based on this evaluation of 30 homes, a community-specific M_{SD} of 0.57 would provide a scientifically valid alternative to the default value of 0.70 in the IEUBK model for the prediction of blood lead concentrations in children in this Texas community.

REFERENCES

Hardy, M.A. 1993. Regression with Dummy Variables. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-093. Newbury Park, CA: Sage.

U.S. EPA (United States Environmental Protection Agency). 1989. Risk Assessment Guidance for Superfund. Volume I. Human Health Evaluation Manual (Part A). Interim Final. EPA/540/1-89/002. Office of Emergency and Remedial Response, Washington, D.C.

U.S. EPA (United States Environmental Protection Agency). 1994. Guidance Manual for the Integrated Exposure Uptake Biokinetic Model for Lead in Children. EPA/540/R-93/081, PB93-963510.

U.S. EPA (United States Environmental Protection Agency). 1998. Short Sheet: IEUBK Model Mass Fraction of Soil in Indoor Dust (M_{SD}) Variable. EPA/540/F/00/008, OSWER Directive 9285.7-34.

U.S. EPA (United States Environmental Protection Agency). 2004. Standard Operating Procedure for Estimating the Soil to Dust Variable. Region 6.

ATTACHMENT A

STATISTICAL OUTPUT
TECHNICAL MEMORANDUM
APRIL 2004